PROCEEDINGS OF TDWG

Biodiversity Information Science and Standards 1: e20320
doi: 10.3897/tdwgproceedings.1.20320

OPEN ACCESS

Conference Abstract

# A tool for collections-specific searches in genetic databases

Michael Trizna ‡

‡ Smithsonian Institution, Washington, DC, United States of America

Corresponding author: Michael Trizna (mike.trizna@gmail.com)

Received: 15 Aug 2017 | Published: 15 Aug 2017

Citation: Trizna M (2017) A tool for collections-specific searches in genetic databases. Proceedings of TDWG 1: e20320. https://doi.org/10.3897/tdwgproceedings.1.20320

## Abstract

It is becoming increasingly important for museums and other scientific collections to quantify the amount of genetic resources being derived from their holdings. Genetic database records, such as GenBank and Barcode of Life (BOLD), have an optional field for indicating the specimen that it derived from, and, on the other side, specimen databases, such as GBIF (gbif.org) and iDigBio (idigbio.org), have an optional field for indicating sequence records that were derived from it. Making connections between the two types of records should be easy, but unfortunately they are made difficult by inconsistent standards. For example, GenBank has a catch-all "country" term that holds all geographic locality data for a specimen, whereas in Darwin Core (DwC) there are 12 atomized levels of locality names.

The software tool described here was originally created for Smithsonian data managers to search genetic databases in a targeted manner for DNA sequences generated from Smithsonian specimens. It is being made open source to be utilized by other scientific institutions to quantify and document the genetic impact of their collections. Other potential uses include checking for data inconsistencies between sequence records and specimen records, and enforcing specimen loan agreements.

© Trizna M. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## Keywords

Collections, GenBank, BOLD, Museums, Biodiversity Informatics, Software, Data Linking

## Presenting author

Michael Trizna